

A Comparative Study of AI-Generated (GPT-4) and Human-crafted MCQs in Programming Education

Year: 2023 | Citations: 101 | Authors: J. Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang

Abstract

There is a constant need for educators to develop and maintain effective up-to-date assessments. While there is a growing body of research in computing education on utilizing large language models (LLMs) in generation and engagement with coding exercises, the use of LLMs for generating programming MCQs has not been extensively explored. We analyzed the capability of GPT-4 to produce multiple-choice questions (MCQs) aligned with specific learning objectives (LOs) from Python programming classes in higher education. Specifically, we developed an LLM-powered (GPT-4) system for generation of MCQs from high-level course context and module-level LOs. We evaluated 651 LLM-generated and 449 human-crafted MCQs aligned to 246 LOs from 6 Python courses. We found that GPT-4 was capable of producing MCQs with clear language, a single correct choice, and high-quality distractors. We also observed that the generated MCQs appeared to be well-aligned with the LOs. Our findings can be leveraged by educators wishing to take advantage of the state-of-the-art generative models to support MCQ authoring efforts.